

Data Mining of Tables: The Barrier for Automation

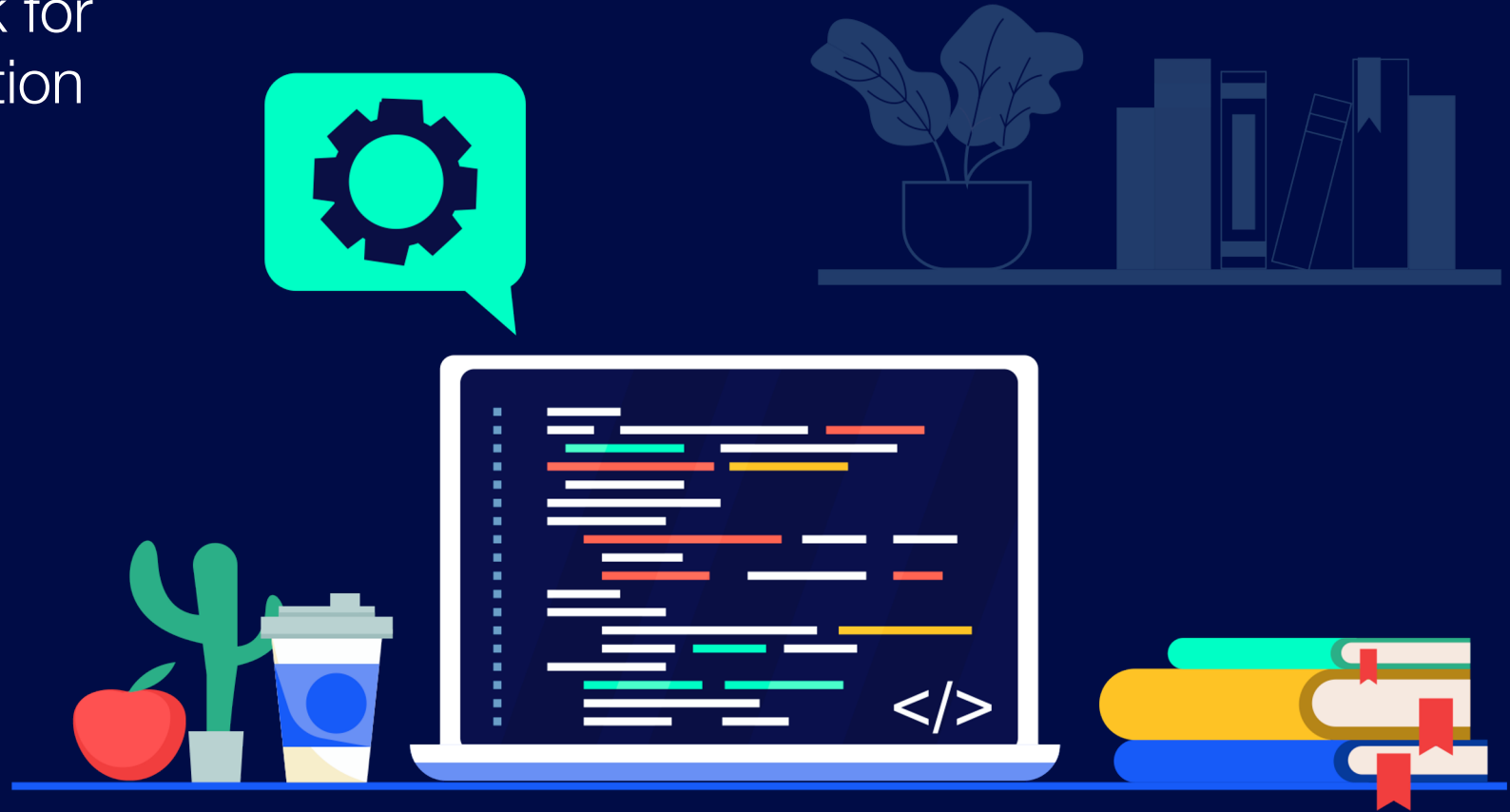


Agenda

- ▶ Introduction
- ▶ Challenge
- ▶ Goals
- ▶ High Level Approach
- ▶ A Deeper Look
- ▶ Examples
- ▶ Use Cases
- ▶ Benefits
- ▶ Q&A

Introduction

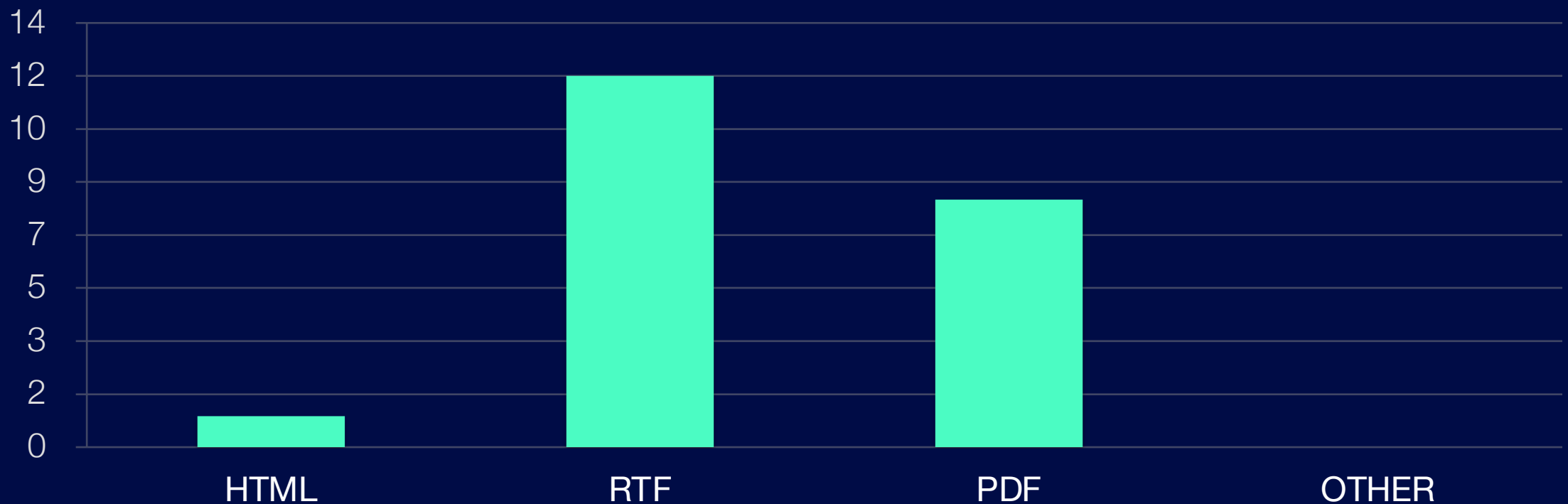
Multi data type stack for
discrepancy localization
within tables



What is the current challenge?

Statistical outputs are in static format, Usually in rtf, html, pdf.

Data hierarchy or structure of the outputs is unknown.



Challenges Using Static Files

1. Lack of standardization.
2. Important metadata is missing.
3. Run analysis on that data.
4. Revision history is missing.
5. Consolidated shared feedback missing.
6. Focal area of communications between the CRO and Sponsor missing.

Table Mining Goals



1

Accurate
& scalable
solution



2

Automatic
seamless process



3

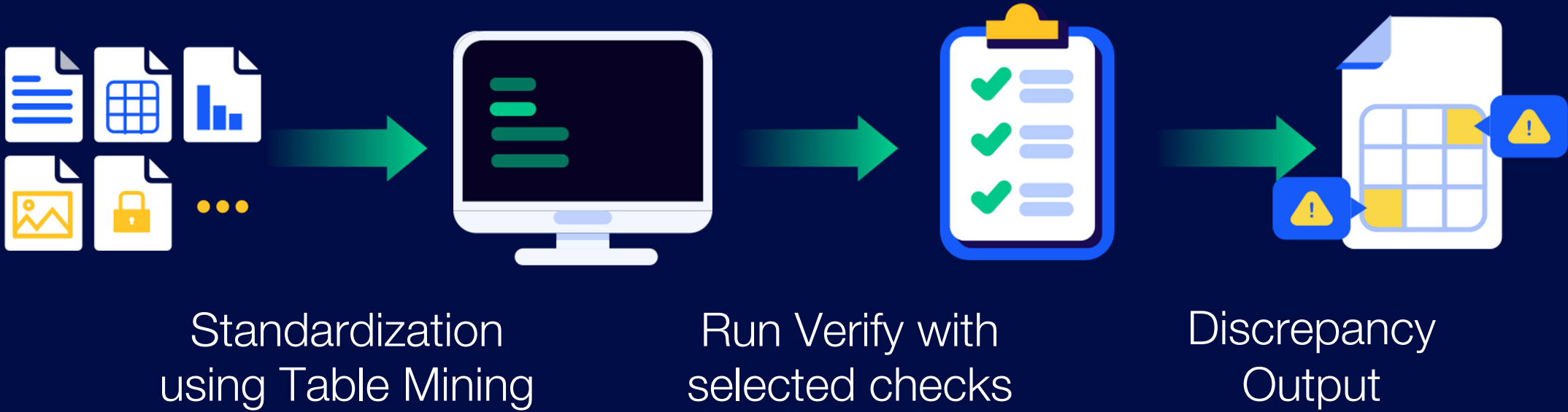
Support different
types of data
format and data
structure



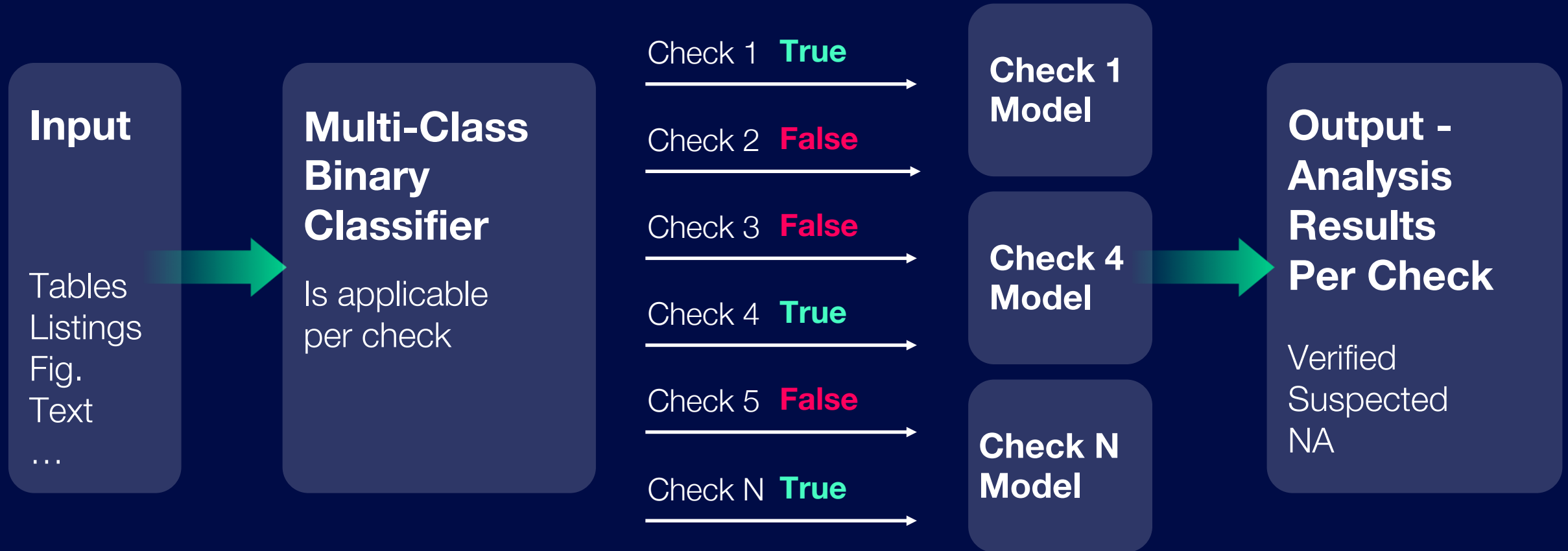
4

Support mixed
data types

High Level Approach



A Deeper Look



Example of Table Mining Output

CSR B456225

Demographics_pp_8w

Table 1.2.3.4

Table 1.2.3.4

Adverse_events_saftey

Baseline_physical_saftey

Baseline_results_saftey

Demographics_completed

Demographics_itt

Demographics_saftey

Disease_4w_1_dose

Disease_4w_itt

Disease_4w_pdd

Disposition

Ecg_baseline_saftey

Table 1.2.3.4
Demographics (Per Protocol Population at 8 weeks)

		Drug A N=48	Drug B N=39	Total N=87
Gender	Male	32 (66.7%)	25 (64.1%)	57 (65.5%)
	Female	16 (33.3%)	14 (35.9%)	30 (34.5%)
Age	<18	0 (0.0%)	0 (0.0%)	0 (0.0%)
	18-30	19 (39.6%)	20 (39.6%)	28 (32.2%)
	31-45	20 (41.7%)	20 (41.7%)	32 (36.8%)
	46-65	9 (18.8%)	9 (18.8%)	27 (31.0%)
	>65	0 (0.0%)	0 (0.0%)	0 (0.0%)
Weight (kg)	N	47	37	84
	Mean	78.9	78.3	87.7

N = 28 (32.2%)

Numeric value: 28

Percentage value: 32.2

Column Header: Total, N = 87

Row Header: Age, 18-30

Table Number: 1.2.3.4

Table Title: Demographics (Per Protocol Population at 8 weeks)

Example of Algo Output

Table 14.1.1.1
Subject Disposition by Site

	Randomized	Treated	Completed 4 weeks of treatment	Completed 8 weeks of treatment	Completed Study	Completed 4 weeks T
Site 001	25	24	24	23	20	1
Site 002	25	25	22	20	19	1
Site 003	24	24	22	22	21	1
Site 004	22	22	19	18	16	1
Site 005	0	0	0	0	0	0
Site 006	18	18	14	13	11	0
Site 007	25	25	22	20	17	1
Site 008	25	25	22	22	20	1
Site 009	24	23	20	20	18	1
Site 010	25	25	22	21	16	1
Site 011	25	25	22	21	17	1
Total	237	236	209	200	175	1

BEACONCURE CONFIDENTIAL SDTM Creation: 08OCT2021 (00:55) Source Data: advs Table Generation: 09OCT2021

USE CASE 1:

Version History of your Deliverables

- Usually each deliverable has several versions and iterations.
- The biostatistician would like to see only the outputs that have changed from previous versions. For example, changes in numeric values between versions

Table 14.1.1.26 Disease Status at 4 weeks (Subjects who received at least one dose)			
Randomized	Drug A (N=119)	Drug B (N=117)	Total (N=236)
Compared with Baseline			
n	119 (100.0%)	117 (100.0%)	236 (100.0%)
Worsening	1 (0.8%)	5 (4.3%)	6 (2.5%)
No Change	28 (23.5%)	29 (24.3%)	62 (26.3%)
Partial Improvement	34 (28.6%)	37 (31.6%)	71 (30.1%)
Complete Healing	56 (47.1%)	41 (35.0%)	97 (41.1%)

The outcome is based on the last visit up to the 4 week visit.
N = number of subjects in the specified group, or the total sample. This value is the denominator for the percentage calculations.
BEACONCURE CONFIDENTIAL SDTM Creation: 08OCT2021 (00:55) Source Data: advs Table Generation: 09OCT2021

USE CASE 2:

Finding Discrepancies in Outputs

healing_rate.rtf

master.rtf

Table 14.1.1.31
Summary of Complete Healing Rate (%) for each population by Treatment (Randomized Population)

Randomized	Drug A (N=118)	Drug B (N=117)	Total (N=235)
	(%)	(%)	(%)
4 weeks ITT	46.7	35.9	40.9
8 weeks ITT	60.0		
4 weeks (>=1 dose)	47.1		
8 weeks (>=1 dose)	60.5		
4 weeks per protocol	57.5		
8 weeks per protocol	67.3		
Subjects 001-005 did not receive any treatment and for the ITT analysis is counted as Worsening. N = number of subjects in the specified group, or the total sample. This value is the denominator for the percentage calculations. BEACONCURE CONFIDENTIAL SDTM Creation: 08OCT2021 (00:55) Source Data: advs Table Generation: 09OCT2021			

Table 14.1.1.2
Analysis Population

	Drug A	Drug B	Total
Randomized Population	120	117	237
Safety Population	119	117	236
Received at least one dose	119	117	236
Completed 4 weeks	89	87	176
Per Protocol Population at 4 weeks	80	61	141
Completed 8 weeks	82	75	157
Per Protocol Population at 8 weeks	48	39	87
Completed Study (12 weeks)	75	69	144
BEACONCURE CONFIDENTIAL SDTM Creation: 08OCT2021 (00:55) Source Data: advs Table Generation: 09OCT2021			

Roadmap

**Increase
Table
Mining**

Generalisation

**Automate
additional manual
admin processes
done by
Biostatisticians**

Benefits

- ✓ Free your statistical analysis outputs.
- ✓ Getting automation of manual processes done quickly.
- ✓ Use it for getting insights on your outputs.
- ✓ Use it as a base for any other analysis application.

Questions?

Schedule a Demo
At booth 27



A large teal triangle graphic pointing to the right, located on the left side of the slide.

Visit us at booth 27

llan@beaconcure.com
www.beaconcure.com